Lecture 4: Heteroskedasticity

Econometric Methods – Warsaw School of Economics

Andrzej Torój

Outline

Outline

Recall: variance-covariance matrix of ε

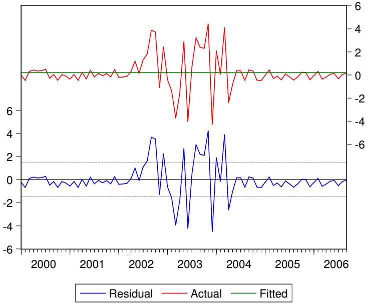
$$E\left(\varepsilon\varepsilon^T\right) =$$

$$= \begin{bmatrix} var\left(\varepsilon_1\right) & cov\left(\varepsilon_1\varepsilon_2\right) & \dots & cov\left(\varepsilon_1\varepsilon_T\right) \\ cov\left(\varepsilon_1\varepsilon_2\right) & var\left(\varepsilon_2\right) & \dots & cov\left(\varepsilon_2\varepsilon_T\right) \\ \vdots & \vdots & \ddots & \vdots \\ cov\left(\varepsilon_1\varepsilon_T\right) & cov\left(\varepsilon_2\varepsilon_T\right) & \dots & var\left(\varepsilon_T\right) \end{bmatrix} =$$

$$\overset{OLS\ assumptions}{=} \begin{bmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{bmatrix}$$

Heteroskedasticity

$$\begin{bmatrix} var(\varepsilon_1) & cov(\varepsilon_1, \varepsilon_2) & \cdots & cov(\varepsilon_1, \varepsilon_T) \\ cov(\varepsilon_1, \varepsilon_2) & var(\varepsilon_2) & \cdots & cov(\varepsilon_2, \varepsilon_T) \\ \vdots & \vdots & \ddots & \\ cov(\varepsilon_1, \varepsilon_T) & cov(\varepsilon_2, \varepsilon_T) & & var(\varepsilon_T) \end{bmatrix} = \sigma^2 \begin{bmatrix} \omega_1 & 0 & \cdots & 0 \\ 0 & \omega_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \\ 0 & 0 & & \omega_T \end{bmatrix}$$



Non-spherical disturbances

variance-covariance matrix of the error term		serial correlation: absent	serial correlation: present
hetero-skedasticity	absent	$\begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \\ 0 & 0 & & \sigma^2 \end{bmatrix}$	$\sigma^2 \begin{bmatrix} 1 & \omega_{12} & \cdots & \omega_{1T} \\ \omega_{12} & 1 & \cdots & \omega_{2T} \\ \vdots & \vdots & \ddots & \\ \omega_{1T} & \omega_{2T} & & 1 \end{bmatrix}$
	present	$\sigma^2 \begin{bmatrix} \omega_1 & 0 & \cdots & 0 \\ 0 & \omega_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \\ 0 & 0 & & \omega_T \end{bmatrix}$	Ω (???)

Consequences of heteroskedasticity

Common features of non-spherical disturbances (see serial correlation):

- no bias, no inconsistency...
- ...but **inefficiency**!

of OLS estimates.

Unlike serial correlation...

heteroskedasticity can occur both in
– **time series data** (e.g. high- and low-volatility periods in financial markets)
– **cross-section data** (e.g. variance of disturbances depends on unit size or some key explanatory variables)

Exercise (1/3)

Credit cards

Based on client-level data, we fit a model that that explains the credit-card-settled expenditures with:

- age;
- income;
- squared income;
- house ownership dummy.

Outline

White test (1)

- Step 1: **OLS regression**
  $y_i = \mathbf{x}_i\boldsymbol{\beta} + \varepsilon_i \qquad \hat{\boldsymbol{\beta}} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y} \qquad \hat{\varepsilon}_i = y_i - \mathbf{x}_i\hat{\boldsymbol{\beta}}$
- Step 2: **auxiliary regression equation**
  $\hat{\varepsilon}_i^2 = \sum_{k,l} x_{k,i}x_{l,i}\beta_{k,l} + v_i$
  E.g. in a model with a constant and 3 regressors $\mathbf{x}_{1t}, \mathbf{x}_{2t}, \mathbf{x}_{3t}$, the auxiliary model contains the following explanatory variables:
  constant, $x_{1t}, x_{2t}, x_{3t}, x_{1t}^2, x_{2t}^2, x_{3t}^2, \underbrace{x_{1t}\cdot x_{2t}, x_{2t}\cdot x_{3t}, x_{1t}\cdot x_{3t}}_{\text{cross terms}}$.
- IDEA: Without heteroskedasticity, the R^2 of the auxiliary equation should be low.

White test (1)

- Step 1: **OLS regression**
  $y_i = \mathbf{x}_i\boldsymbol{\beta} + \varepsilon_i \qquad \hat{\boldsymbol{\beta}} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y} \qquad \hat{\varepsilon}_i = y_i - \mathbf{x}_i\hat{\boldsymbol{\beta}}$
- Step 2: **auxiliary regression equation**
  $\hat{\varepsilon}_i^2 = \sum_{k,l} x_{k,i}x_{l,i}\beta_{k,l} + v_i$
  E.g. in a model with a constant and 3 regressors $\mathbf{x}_{1t}, \mathbf{x}_{2t}, \mathbf{x}_{3t}$, the auxiliary model contains the following explanatory variables:
  constant, $x_{1t}, x_{2t}, x_{3t}, x_{1t}^2, x_{2t}^2, x_{3t}^2, \underbrace{x_{1t}\cdot x_{2t}, x_{2t}\cdot x_{3t}, x_{1t}\cdot x_{3t}}_{\textit{cross terms}}$.
- IDEA: Without heteroskedasticity, the R^2 of the auxiliary equation should be low.

White test (1)

- Step 1: **OLS regression**
  $y_i = \mathbf{x}_i\boldsymbol{\beta} + \varepsilon_i \qquad \hat{\boldsymbol{\beta}} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y} \qquad \hat{\varepsilon}_i = y_i - \mathbf{x}_i\hat{\boldsymbol{\beta}}$
- Step 2: **auxiliary regression equation**
  $\hat{\varepsilon}_i^2 = \sum_{k,l} x_{k,i}x_{l,i}\beta_{k,l} + v_i$
  E.g. in a model with a constant and 3 regressors $\mathbf{x}_{1t}, \mathbf{x}_{2t}, \mathbf{x}_{3t}$, the auxiliary model contains the following explanatory variables:
  constant, $x_{1t}, x_{2t}, x_{3t}, x_{1t}^2, x_{2t}^2, x_{3t}^2, \underbrace{x_{1t}\cdot x_{2t}, x_{2t}\cdot x_{3t}, x_{1t}\cdot x_{3t}}_{\textit{cross terms}}.$
- IDEA: Without heteroskedasticity, the R^2 of the auxiliary equation should be low.

White test (2)

White test

H_0 : heteroskedasticity absent
H_1 : heteroskedasticity present
$W = TR^2 \sim \chi^2(k^*)$
k^* – number of explanatory variables in the auxiliary regresson (excluding constant)
CAUTION! It's a **weak** test (i.e. low power to reject the null)

Goldfeld-Quandt test

- split the sample (T observations) into 2 subsamples ($T = n_1 + n_2$)
- test for equality of error term variance in both subsamples

Goldfeld-Quandt

$H_0 : \sigma_1^2 = \sigma_2^2$ equal variance in both subsamples (homoskedasticity)
$H_1 : \sigma_1^2 > \sigma_2^2$ higher variance in the subsample indexed as 1

$$F(n_1 - k, n_2 - k) = \frac{\sum_{i=1}^{n_1} \hat{\varepsilon}_i^2 / (n_1 - k)}{\sum_{i=n_1+1}^{T} \hat{\varepsilon}_i^2 / (n_2 - k)}$$

CAUTION! This makes sense only when we index the subsample with higher variance as 1. Otherwise we never reject H_0.

Breusch-Pagan test

- variance of the disturbances can be explained with a variable set contained in matrix **Z** (like explanatory variables for **y** in the matrix **X**)

Breusch-Pagan test

H_0 : homoskedasticity
H_1 : heteroskedasticity

$$BP = \frac{\left(\hat{\boldsymbol{\varepsilon}^2} - \hat{\sigma}^2 \mathbf{1}\right)^T \mathbf{Z}\left(\mathbf{Z}^T\mathbf{Z}\right)^{-1}\mathbf{Z}^T\left(\hat{\boldsymbol{\varepsilon}^2} - \hat{\sigma}^2 \mathbf{1}\right)}{\sum_{i=1}^{n} \hat{\varepsilon}_i^2/(n-k)}$$

where $\hat{\boldsymbol{\varepsilon}^2} = \begin{bmatrix} \varepsilon_1^2 & \varepsilon_2^2 & \ldots & \varepsilon_T^2 \end{bmatrix}^T$, $\mathbf{1} = \begin{bmatrix} 1 & 1 & \ldots & 1 \end{bmatrix}^T$. The test statistic is χ^2-distributed with degrees of freedom equal to the number of regressors in the matrix **Z**.

Exercise (2/3)

Credit cards

1. Does the White test detect heteroskedasticity?
2. Split the sample into two equal subsamples: high-income and low-income. Check if the variance differs between the two sub-samples. (You need to sort the data and restrict the sample to a sub-sample twice, each time calculating the appropriate statistics.)
3. Perform the Breusch-Pagan test, assuming that the variance depends only on the income and squared income (and a constant).

Outline

White robust SE

- unlike Newey-West robust SE (robust to both serial correlation and heteroskedasticity), **White**'s SE robust **only to heteroskedasticity** (the former were proposed later and generalized White's work)
- *White (1980)*:
$$Var\left(\hat{\beta}\right) = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\left(\sum_{t=1}^{T}\hat{\varepsilon}_t^2\mathbf{x}_t\mathbf{x}_t^T\right)\left(\mathbf{X}^T\mathbf{X}\right)^{-1}$$
- they share the same features as Newey-West SE (see: serial correlation), i.e. correct statistical inference without improving estimation efficiency of the parameters themselves

White robust SE

- unlike Newey-West robust SE (robust to both serial correlation and heteroskedasticity), **White**'s SE robust **only to heteroskedasticity** (the former were proposed later and generalized White's work)
- ***White (1980)***:

$$Var\left(\hat{\beta}\right) = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\left(\sum_{t=1}^{T}\hat{\varepsilon}_t^2\mathbf{x}_t\mathbf{x}_t^T\right)\left(\mathbf{X}^T\mathbf{X}\right)^{-1}$$

- they share the same features as Newey-West SE (see: serial correlation), i.e. correct statistical inference without improving estimation efficiency of the parameters themselves

White robust SE

- unlike Newey-West robust SE (robust to both serial correlation and heteroskedasticity), **White**'s SE robust **only to heteroskedasticity** (the former were proposed later and generalized White's work)
- ***White (1980)***:
$$Var\left(\hat{\beta}\right) = \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\left(\sum_{t=1}^{T}\hat{\varepsilon}_t^2\mathbf{x}_t\mathbf{x}_t^T\right)\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}$$
- they share the same features as Newey-West SE (see: serial correlation), i.e. correct statistical inference without improving estimation efficiency of the parameters themselves

Weighted Least Squares estimator (1)

- $\mathbf{y} = \mathbf{X}\beta + \varepsilon \qquad \varepsilon \sim \left(E\left[\varepsilon\right] = \mathbf{0}, E\left[\varepsilon\varepsilon^T\right] = \mathbf{\Omega}\right)$
- Recall the GLS estimator. Under heteroskedasticity, we know that the variance-covariance matrix $\mathbf{\Omega} = \begin{bmatrix} \omega_1 & 0 & \dots & 0 \\ 0 & \omega_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \omega_T \end{bmatrix}$, which implies $\mathbf{\Omega}^{-1} = \begin{bmatrix} \frac{1}{\omega_1} & 0 & \dots & 0 \\ 0 & \frac{1}{\omega_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{\omega_T} \end{bmatrix}$.

Weighted Least Squares estimator (1)

- $\mathbf{y} = \mathbf{X}\beta + \varepsilon \qquad \varepsilon \sim \left(E\left[\varepsilon\right] = \mathbf{0}, E\left[\varepsilon\varepsilon^T\right] = \mathbf{\Omega}\right)$
- Recall the GLS estimator. Under heteroskedasticity, we know that the variance-covariance matrix $\mathbf{\Omega} = \begin{bmatrix} \omega_1 & 0 & \dots & 0 \\ 0 & \omega_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \omega_T \end{bmatrix}$, which implies $\mathbf{\Omega}^{-1} = \begin{bmatrix} \frac{1}{\omega_1} & 0 & \dots & 0 \\ 0 & \frac{1}{\omega_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{\omega_T} \end{bmatrix}$.

Weighted Least Squares estimator (2)

- Knowing the vector $\begin{bmatrix} \frac{1}{\omega_1} & \frac{1}{\omega_2} & \dots & \frac{1}{\omega_T} \end{bmatrix}$ we can immediately apply GLS:

$$
\begin{aligned}
\hat{\beta}^{WLS} &= [\mathbf{X}^T\boldsymbol{\Omega}^{-1}\mathbf{X}]^{-1}\mathbf{X}^T\boldsymbol{\Omega}^{-1}\mathbf{y} = \\
&= \left(\begin{bmatrix} \boldsymbol{x}_1^T & \boldsymbol{x}_2^T & \dots & \boldsymbol{x}_T^T \end{bmatrix} \begin{bmatrix} \frac{1}{\omega_1} & 0 & \dots & 0 \\ 0 & \frac{1}{\omega_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{\omega_T} \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \\ \vdots \\ \boldsymbol{x}_T \end{bmatrix} \right)^{-1} \cdot \\
&\quad \cdot \begin{bmatrix} \boldsymbol{x}_1^T & \boldsymbol{x}_2^T & \dots & \boldsymbol{x}_T^T \end{bmatrix} \begin{bmatrix} \frac{1}{\omega_1} & 0 & \dots & 0 \\ 0 & \frac{1}{\omega_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{\omega_T} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{bmatrix} = \\
&= \left(\sum_{i=1}^{T} \frac{1}{\omega_i} \boldsymbol{x}_i^T \boldsymbol{x}_i \right)^{-1} \left(\sum_{i=1}^{T} \frac{1}{\omega_i} \boldsymbol{x}_i^T y_i \right)
\end{aligned}
$$

- This vector can hence be interpreted as a **vector of weights**, associated with individual observations in the estimation (hence: "weighted" least squares).

Weighted Least Squares estimator

- $$\hat{\beta}^{WLS} = \left(\sum_{i=1}^{T}\frac{1}{\omega_i}\mathbf{x}_i^T\mathbf{x}_i\right)^{-1}\left(\sum_{i=1}^{n}\frac{1}{\omega_i}\mathbf{x}_i^T y_i\right) = \left(\sum_{i=1}^{n}\frac{\mathrm{x}_i}{\sqrt{\omega_i}}\frac{\mathrm{x}_i^T}{\sqrt{\omega_i}}\right)^{-1}\left(\sum_{i=1}^{T}\frac{\mathrm{x}_i^T}{\sqrt{\omega_i}}\frac{y_i}{\sqrt{\omega_i}}\right)$$
- The WLS estimation is hence equivalent to OLS estimation using data transformed in the following way:
$$\mathbf{y}^* = \begin{bmatrix} y_1/\sqrt{\omega_1} \\ y_2/\sqrt{\omega_2} \\ \vdots \\ y_T/\sqrt{\omega_T} \end{bmatrix} \qquad \mathbf{X}^* = \begin{bmatrix} \mathrm{x}_1/\sqrt{\omega_1} \\ \mathrm{x}_2/\sqrt{\omega_2} \\ \vdots \\ \mathrm{x}_T/\sqrt{\omega_T} \end{bmatrix}$$

Conclusion

Weights for individual observations are the inverse of the disturbance variance in individual periods. Under OLS, these weights are a unit vector.

Weighted Least Squares estimator

- $$\begin{aligned}\hat{\beta}^{WLS} &= \left(\sum_{i=1}^{T}\frac{1}{\omega_i}\mathbf{x_i}^T\mathbf{x_i}\right)^{-1}\left(\sum_{i=1}^{n}\frac{1}{\omega_i}\mathbf{x_i}^T y_i\right) = \\ &= \left(\sum_{i=1}^{n}\frac{\mathbf{x}_i}{\sqrt{\omega_i}}\frac{\mathbf{x}_i^T}{\sqrt{\omega_i}}\right)^{-1}\left(\sum_{i=1}^{T}\frac{\mathbf{x}_i^T}{\sqrt{\omega_i}}\frac{y_i}{\sqrt{\omega_i}}\right)\end{aligned}$$

- The WLS estimation is hence equivalent to OLS estimation using data transformed in the following way:

$$\mathbf{y}^* = \begin{bmatrix} y_1/\sqrt{\omega_1} \\ y_2/\sqrt{\omega_2} \\ \vdots \\ y_T/\sqrt{\omega_T} \end{bmatrix} \qquad \mathbf{X}^* = \begin{bmatrix} \mathbf{x}_1/\sqrt{\omega_1} \\ \mathbf{x}_2/\sqrt{\omega_2} \\ \vdots \\ \mathbf{x_T}/\sqrt{\omega_T} \end{bmatrix}$$

Conclusion

Weights for individual observations are the inverse of the disturbance variance in individual periods. Under OLS, these weights are a unit vector.

- $$\begin{aligned}\hat{\beta}^{WLS} &= \left(\sum_{i=1}^{T} \frac{1}{\omega_i} \mathbf{x}_i^T \mathbf{x}_i\right)^{-1} \left(\sum_{i=1}^{n} \frac{1}{\omega_i} \mathbf{x}_i^T y_i\right) = \\ &= \left(\sum_{i=1}^{n} \frac{\mathbf{x}_i}{\sqrt{\omega_i}} \frac{\mathbf{x}_i^T}{\sqrt{\omega_i}}\right)^{-1} \left(\sum_{i=1}^{T} \frac{\mathbf{x}_i^T}{\sqrt{\omega_i}} \frac{y_i}{\sqrt{\omega_i}}\right)\end{aligned}$$

- The WLS estimation is hence equivalent to OLS estimation using data transformed in the following way:

$$\mathbf{y}^* = \begin{bmatrix} y_1/\sqrt{\omega_1} \\ y_2/\sqrt{\omega_2} \\ \vdots \\ y_T/\sqrt{\omega_T} \end{bmatrix} \qquad \mathbf{X}^* = \begin{bmatrix} \mathbf{x}_1/\sqrt{\omega_1} \\ \mathbf{x}_2/\sqrt{\omega_2} \\ \vdots \\ \mathbf{x_T}/\sqrt{\omega_T} \end{bmatrix}$$

Conclusion

Weights for individual observations are the inverse of the disturbance variance in individual periods. Under OLS, these weights are a unit vector.

How to find ω_1, ω_2, ..., ω_T?

Unknown and, with T observations, cannot be estimated. The most popular solutions include:

- **Way 1:**
  - Split the sample into subsamples.
  - Estimate the model in each subsample via OLS to obtain the vector $\hat{\varepsilon}$.
  - In every subsample i estimate the variance of error terms $\hat{\sigma}_i^2$.
  - Assign the weight $\frac{1}{\hat{\sigma}_i^2}$ to all the observations in the subsample i.

How to find ω_1, ω_2, ..., ω_T?

Unknown and, with T observations, cannot be estimated. The most popular solutions include:

- **Way 1:**
  - Split the sample into subsamples.
  - Estimate the model in each subsample via OLS to obtain the vector $\hat{\varepsilon}$.
  - In every subsample i estimate the variance of error terms $\hat{\sigma}_i^2$.
  - Assign the weight $\frac{1}{\hat{\sigma}_i^2}$ to all the observations in the subsample i.

How to find ω_1, ω_2, ..., ω_T?

Unknown and, with T observations, cannot be estimated. The most popular solutions include:

- **Way 1:**
  - Split the sample into subsamples.
  - Estimate the model in each subsample via OLS to obtain the vector $\hat{\varepsilon}$.
  - In every subsample i estimate the variance of error terms $\hat{\sigma}_i^2$.
  - Assign the weight $\frac{1}{\hat{\sigma}_i^2}$ to all the observations in the subsample i.

How to find ω_1, ω_2, ..., ω_T?

Unknown and, with T observations, cannot be estimated. The most popular solutions include:

- **Way 1:**
    - Split the sample into subsamples.
    - Estimate the model in each subsample via OLS to obtain the vector $\hat{\varepsilon}$.
    - In every subsample i estimate the variance of error terms $\hat{\sigma}_i^2$.
    - Assign the weight $\frac{1}{\hat{\sigma}_i^2}$ to all the observations in the subsample i.

How to find ω_1, ω_2, ..., ω_T?

Unknown and, with T observations, cannot be estimated. The most popular solutions include:

- **Way 1:**
  - Split the sample into subsamples.
  - Estimate the model in each subsample via OLS to obtain the vector $\hat{\varepsilon}$.
  - In every subsample i estimate the variance of error terms $\hat{\sigma}_i^2$.
  - Assign the weight $\frac{1}{\hat{\sigma}_i^2}$ to all the observations in the subsample i.

- **Way 2:**
  - Estimate the model via OLS to obtain the vector $\hat{\varepsilon}$.
  - Regress $\hat{\varepsilon}_t^2$ against a set of potential explanatory variables (when done automatically, usually all the regressors from the base equation plus possibly their squares).
  - Take theoretical value of $\hat{\varepsilon}_t^2$ from this regression – say, e_t^2 – as a proxy of variance (one cannot use $\hat{\varepsilon}_t^2$ itself, as it does not measure variance adequately – it is just one draw from a distribution, while the theoretical value summarizes a number of draws made under similar conditions regarding the explanatory variables for variance).
  - Use $\frac{1}{e_t^2}$ as weights for individual observations t.
  - In practice, it is common to regress $ln\left(\hat{\varepsilon}_t^2\right)$ rather than $\hat{\varepsilon}_t^2$. In this way we compute $\widehat{\left(\varepsilon_t^2\right)} = exp\left(\widehat{ln\left(\varepsilon_t^2\right)}\right)$ which blocks negative values of error terms' variance.

- **Way 2:**
  - Estimate the model via OLS to obtain the vector $\hat{\varepsilon}$.
  - Regress $\hat{\varepsilon}_t^2$ against a set of potential explanatory variables (when done automatically, usually all the regressors from the base equation plus possibly their squares).
  - Take theoretical value of $\hat{\varepsilon}_t^2$ from this regression – say, e_t^2 – as a proxy of variance (one cannot use $\hat{\varepsilon}_t^2$ itself, as it does not measure variance adequately – it is just one draw from a distribution, while the theoretical value summarizes a number of draws made under similar conditions regarding the explanatory variables for variance).
  - Use $\frac{1}{e_t^2}$ as weights for individual observations t.
  - In practice, it is common to regress $ln\left(\hat{\varepsilon}_t^2\right)$ rather than $\hat{\varepsilon}_t^2$. In this way we compute $\widehat{\left(\varepsilon_t^2\right)} = exp\left(ln\widehat{\left(\varepsilon_t^2\right)}\right)$ which blocks negative values of error terms' variance.

- **Way 2:**
  - Estimate the model via OLS to obtain the vector $\hat{\varepsilon}$.
  - Regress $\hat{\varepsilon}_t^2$ against a set of potential explanatory variables (when done automatically, usually all the regressors from the base equation plus possibly their squares).
  - Take theoretical value of $\hat{\varepsilon}_t^2$ from this regression – say, e_t^2 – as a proxy of variance (one cannot use $\hat{\varepsilon}_t^2$ itself, as it does not measure variance adequately – it is just one draw from a distribution, while the theoretical value summarizes a number of draws made under similar conditions regarding the explanatory variables for variance).
  - Use $\frac{1}{e_t^2}$ as weights for individual observations t.
  - In practice, it is common to regress $ln\left(\hat{\varepsilon}_t^2\right)$ rather than $\hat{\varepsilon}_t^2$. In this way we compute $\left(\widehat{\varepsilon_t^2}\right) = exp\left(ln\left(\widehat{\varepsilon_t^2}\right)\right)$ which blocks negative values of error terms' variance.

- **Way 2:**
  - Estimate the model via OLS to obtain the vector $\hat{\varepsilon}$.
  - Regress $\hat{\varepsilon}_t^2$ against a set of potential explanatory variables (when done automatically, usually all the regressors from the base equation plus possibly their squares).
  - Take theoretical value of $\hat{\varepsilon}_t^2$ from this regression – say, e_t^2 – as a proxy of variance (one cannot use $\hat{\varepsilon}_t^2$ itself, as it does not measure variance adequately – it is just one draw from a distribution, while the theoretical value summarizes a number of draws made under similar conditions regarding the explanatory variables for variance).
  - Use $\frac{1}{e_t^2}$ as weights for individual observations t.
  - In practice, it is common to regress $ln\left(\hat{\varepsilon}_t^2\right)$ rather than $\hat{\varepsilon}_t^2$. In this way we compute $\left(\widehat{\varepsilon_t^2}\right) = exp\left(ln\left(\widehat{\varepsilon_t^2}\right)\right)$ which blocks negative values of error terms' variance.

- **Way 2:**
  - Estimate the model via OLS to obtain the vector $\hat{\varepsilon}$.
  - Regress $\hat{\varepsilon}_t^2$ against a set of potential explanatory variables (when done automatically, usually all the regressors from the base equation plus possibly their squares).
  - Take theoretical value of $\hat{\varepsilon}_t^2$ from this regression – say, e_t^2 – as a proxy of variance (one cannot use $\hat{\varepsilon}_t^2$ itself, as it does not measure variance adequately – it is just one draw from a distribution, while the theoretical value summarizes a number of draws made under similar conditions regarding the explanatory variables for variance).
  - Use $\frac{1}{e_t^2}$ as weights for individual observations t.
  - In practice, it is common to regress $ln\left(\hat{\varepsilon}_t^2\right)$ rather than $\hat{\varepsilon}_t^2$. In this way we compute $\widehat{\left(\varepsilon_t^2\right)} = exp\left(\widehat{ln\left(\varepsilon_t^2\right)}\right)$ which blocks negative values of error terms' variance.

Exercise (3/3)

Credit cards

1. Split the sample into two equal subsamples and use WLS (way 1).
2. Use all the explanatory variables and their sqares as the regressors in the variance equation and use WLS (way 2).
3. Compare the parameter values between OLS and the two variants of WLS;
4. Compare variable significance between OLS, OLS with White's robust standard errors and the two variants of WLS.

Readings

- Greene: chapter "Generalized Regression Model and Heteroscedasticity"

Homework

Gasoline demand model

Verify the presence of heteroskedasticity in the model considered in the previous lecture.

Programming

Write an R-function performing an automated heteroskedasticity correction a la "way 2" in this presentation, using WLS, and using all the right-hand side variables as potential determinants of error term variance.